Genotyping Technologies for Genetic Research

Jiannis Ragoussis

Genomics Laboratory, Wellcome Trust Centre for Human Genetics, Oxford University, Oxford OX3 7BN, United Kingdom; email: ioannis.ragoussis@well.ox.ac.uk

Annu. Rev. Genomics Hum. Genet. 2009. 10:117–33

First published online as a Review in Advance on May 19, 2009

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

This article's doi: 10.1146/annurev-genom-082908-150116

1527-8204/09/0922-0117$20.00

Key Words

whole-genome association, GeneChip, Beadchip, P CNV, molecular haplotyping

Abstract

The past few years have seen enormous advances in genotyping technology, including chips that accommodate in excess of 1 million SNP assays. In addition, the cost per genotype has been driven down to levels unimagined only a few years ago. These developments have resulted in an explosion of positive whole-genome association studies and the identification of many new genes for common diseases. Here I review high-throughput genotyping platforms as well as other approaches for lower numbers of assays but high sample throughput, which play an important role in genotype validation and study replication. Further, the utility of SNP arrays for detecting structural variation through the development of genotyping algorithms is reviewed and methods for long-range haplotyping are presented. It is anticipated that in the future, sample throughput and cost savings will be increased further through the combination of automation, microfluidics, and nanotechnologies.

INTRODUCTION

The sequencing of the human genome in 2001 (51) enabled positional cloning of genes for monogenic diseases to be carried out efficiently and with high degree of success. Over 1500 genes were identified, but success was mainly limited to rare diseases caused by genetic mutations. Common, complex diseases required in most cases a whole-genome association approach using very high numbers (over 100,000) of genetic markers and thousands of samples. As these markers had to be identified and validated first, the HapMap project was launched with the aim of capturing genetic variation (77) in the form of single-nucleotide polymorphisms (SNPs) and providing haplotype information for a representative sample derived from European, African, and Asian populations. The second phase of this project, now complete, has produced over 3.1 million validated SNPs (2), although the actual number is thought to be in the region of 10 million.

The HapMap (**http://www.hapmap.org/**) and SNP databases (dbSNP; provided valuable information for association study design and data interpretation, as well as information for SNP assay designs. In combination with technological developments the genomic variation information in databases was used by companies to generate cost-efficient, high-throughput genotyping products, suitable for linkage and whole-genome association (WGA) studies. The enabling technologies were combined with well-designed studies aiming to meet strict criteria for establishing an initial association report and a positive replication set in (12). Points relevant for the technological developments were that (*a*) assay call rates, error rates (estimated using internal or external duplicates), concordance with published data (for example using HapMap samples), Mendelian consistency and deviations from the Hardy-Weinberg equilibrium should be reported; (*b*) a subset of notable SNPs should be evaluated using a different genotyping platform and (c) positive findings should be validated in independent, adequately powered replication studies.

Within a few years from conception, WGA studies have proved successful in identifying genes involved in many common diseases. Up to November 2008 at least 209 reports (**http://www.genome.gov/gwastudies/**) of successful WGA studies had been published, including those to identify genes for such common diseases as asthma (63) and diabetes (81, 85, 103). The WTCC Consortium, which presented the largest study of seven common diseases (1), identified susceptibility genes for bipolar disorder, coronary heart disease, Crohn's disease, type one and type two diabetes and rheumatoid arthritis.

In the meantime, structural variation has emerged as a significant contributor to human genetic variation in addition to sequence variants (76). Copy number variants (CNVs) are defined as fragments of the genome that are larger than 1 kb and vary in copy number between individuals (25). CNVs are now investigated for their contribution to common diseases such as autism (57, 83; reviewed in 23), prompting a new race to incorporate assays for CNVs within the SNP genotyping chips and new analysis algorithms to infer CNVs from SNP genotyping data. A further challenge has been to cover the regions of the genome deemed "unSNPable" owing to recombination hotspots and CNVs (28).

In parallel are endeavors to sequence the genome of model organisms and identify SNPs to enable as equally efficient genetic studies as in humans; analyses of 44 different genomes are currently under way, summarized in db SNP build 129 (April 2008) at NCBI.

The emergence of reliable and ultra-high-throughput genotyping platforms, able to assay for 1 million SNPs or more, has played a decisive role in the success of WGA studies. These platforms were supplemented by technologies that allowed rapid and cost-efficient genotyping of smaller marker sets (10-20,000) in replication cohorts often larger than the ones used for the original study. For example, Todd et al. (94) followed up the 2000 cases and 3000 controls WTCCC type 1 diabetes study with a

Haplotype Map (HapMap): the depiction of over 3.1 million SNP locations on the genome, compiled using 269 samples from four human populations (27). This is an international project involving a consortium of nine research groups from five countries

Single-nucleotide polymorphisms (SNP): these represent common (defined as >1%) variation between individuals in a population

Whole-genome association or genome-wide association (WGA or GWA): when a large number of markers (typically 100,000 or more in human) are analyzed for association with a disease or trait

Copy number variants (CNVs): segments of DNA with a different number of copies when compared to the reference genome, typically deletions or duplications

replication involving 5,000 controls, 4,000 cases and 2,997 trios using 11 SNPs. At the same time, robust assays have also been developed to enable the cost-efficient application of SNP genotyping in clinical environments.

In this review, technologies significant in genetic research are presented, and their relative strengths and weaknesses are discussed. These technologies can be divided into systems used for WGA studies (i.e., 100,000-1 mi SNP assays: Affymetrix GeneChip; Illumina Infinium Beadchips, Perlegen, and Invader); and systems for replication and validation of findings, linkage analysis, or candidate gene approaches (1–100,000 SNP assays; Affymetrix GeneChip and MIP, Illumina Goldengate, and Infinium assays, Invader, Sequenom MassARRAY; SNPlex, SNP stream, Taqman, and Centaurus assays).

TECHNOLOGIES SUITABLE FOR WHOLE-GENOME ASSOCIATION STUDIES

For human genome-wide association studies (GWAS), assays for at least 100,000 SNPs are needed in the initial stage. To date, four such technologies have been used: Invader assays, The Perlegen Genotyping Platform, Affymetrix GeneChips, and Illumina's Infinium Beadchips. The Affymetrix and Illumina chips offer the highest marker densities with 1.8 and 1.2 Mi assays, respectively, but in general, although all aim to cover common variation, the SNP assays are selected and fixed based on criteria set by the individual companies. As a result, the high-density SNPs offer limited opportunities for individual customers to dictate the content.

Illumina's Infinium Beadchips

The Infinium assays started in the form of Infinium I, designed for 10,000–100,000 multiplexed assays, whereas the later introduced Infinium II can be used for up to 1 million assays. Both assays include first a whole-genome amplification step, followed by hybridization to bead arrays of 50-bp-long capture probes. In the Infinium I assay (30), the locus-specific sequences include an allele-specific 3′ terminal base. An allele-specific primer extension reaction is used to incorporate biotin-labeled nucleotides for positive detection. In the Infinium II assay (87), the beads contain locus-specific probes, and allelic discrimination occurs through single-base primer extension reactions using (**Figure 1**). The two-color system in Infinium II restricts somewhat the classes of SNPs that can be genotyped by not including A/T and C/G SNPs. High pass rates and accuracy (>99.9) are performance characteristics for the Infinium assays.

At present, Illumina offers Beadchips with 300 k (cytoSNP-12), 370 k, 510 k, 660 k, and 1 mi assays. Illumina has chosen to base the SNP selection strategy primarily on the HapMap project results and to supplement with genic and nonsynonymous SNPs plus assays for CNVs and unSNPable regions (28). The SNPs are therefore selected to tag haplotypes, and in the case of the 1 mi chip, for all human populations analyzed by the HapMap project. The genome coverage at $r^2 = 0.8$ (r^2 signifies the pairwise correlation between a SNP used and a potentially captured SNP (4)) is projected to

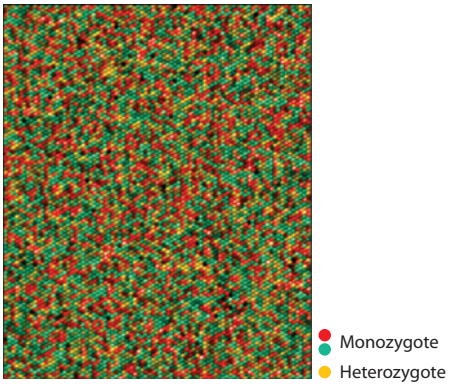

Figure 1

Section derived from an Illumina 610S Quad Beadchip under standard scanning conditions. The beads are 2 μm in diameter and carry the single locus-specific probe. After a SBE reaction the beads are labeled either green or red for the monozygote state or yellow for the heterozygote.

range from 0.76 for the Yoruban population to 0.95 for Europeans. In total, the chip contains 1,199,187 assays, including 21,877 nonsynonymous SNPs, 340,585 assays targeting CNVs, and 38,619 assays for the unSNPable genome.

The DNA requirements are low, ranging from 200 ng for the 300 k–660 k chip to 400 ng for the 1 mi chip.

An advantage of the Beadchips is that up to 60,800 custom assays can be added to supplement the standard assays. Custom assay conversion rates were predicted to reach a possible 89.9% (87). (Details on this are discussed in the next section.)

The system consists of a high-resolution scanner (Beadstation), thermoblocks, and a hybridization oven. Liquid handling is not necessary but is required for high-sample throughput and is an option (provided by Illumina) for setting up chip hybridization and washes.

One important issue is the development of a genotyping algorithm for the Infinium assays. Illumina offers the GeneCall algorithm, which is based on prior training datasets. Two further algorithms have been developed: Illuminus (92), a model-based approach that pools information across individual sample data to achieve improved call rates compared to GeneCall, and GenoSNP (29), which uses within-sample information and is thus suitable for achieving high call rates even when the number of project samples is very small or when the samples are derived from a population that is not well characterized. Illuminus has been tested and found to perform well when whole-genome amplified DNA is used (91).

Affymetrix GeneChip Human Mapping arrays

The GeneChip assays are based on allelic discrimination by direct hybridization of genomic DNA to arrays containing locus- and allele-specific oligonucleotides (25 mers). These oligonucleotides represent either perfect match (PM) or mismatch (MM) probes to each SNP. Initially, 12 PM and MM probes were used for each SNP assay. The MM probes are required to measure background, but recently Affymetrix has applied only 2 PM probes per SNP in quadruplicates in the version 5.0 arrays or higher, thus achieving a very high assay number per chip. The procedure allows the detection of 10,000–2,000,000 SNPs. For the GeneChip assays to work efficiently, the complexity of the genomic DNA must be reduced through digestion with restriction endonucleases and fractionation (42). Genomic DNA is digested with restriction enzymes appropriate for the number of SNPs to be interrogated (i.e., Xba I for the 10 K array, Xba I and Hind III for the 100 K assays, Sty I and Nsp I for the 500 K or higher assay) and the 400–800-bp range is used to ligate adaptors. Following a PCR amplification step, the products are end-labeled and hybridized (**Figure 2**). The highest density arrays at present are the 5.0 and 6.0 versions containing very small 5 μm features, taking advantage of development in scanner technology. The 6.0 chip contains over 1.8 million markers consisting of 906,600 polymorphic and 946,000 nonpolymorphic markers, the latter targeting mainly CNV regions. The 6.0 array reaches a median value of 680 bp for intermarker distance. Data for the GeneChip SNP arrays indicate pass rates of >95% and accuracy of >99% (58). A dynamic

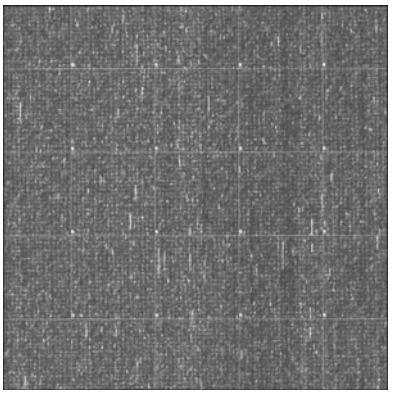

Figure 2

Scanning image of an Affymetrix 250 k GeneChip, Nsp I digest assays. The GeneChip features are 5 μm and the assay is single color. The array carries perfect match probes for each allele at a locus and mismatch probes for determining background.

model-based algorithm (DM) was introduced by Affymetrix (20) that is able to accurately call over 100,000 genotypes. Recently, several algorithms have been developed by Affymetrix (BRLMM in 2006 and Birdseed in 2007), while the Wellcome Trust Case Control Consortium (WTCCC) developed the software CHIAMO (1) based on data derived from large sample collections, and all leading to high call rates >99%. Apart from GWA studies, the GeneChip genotyping system is suitable for detecting copy number changes (100) and loss of heterozygosity, and therefore suitable for clinical applications, for example, in tumor classification (55).

The equipment required for the Affymetrix GeneChip is based on the GeneChip 3000 scanner, workstation, fluidics station, and hybridization oven. Several fluidics stations are required; a LIMS system is also available. The complete system, although expensive, can be used for a number of expression profiling or resequencing applications (73).

Perlegen's Genotyping Method

Perlegen has developed genotyping assays based on their own designs and chips made by Affymetrix. Perlegen uses not only PM and MM probes similar to Affymetrix, but tiles the probes across the SNP position in both forward and reverse strand orientations. The position interrogating the SNP also moves within the 25-bp probe by 4 bp between positions 11 and 15 (36). The target DNA is prepared through a series of long-range PCR reactions that amplify the loci containing the SNPs from each individual sample. The PCR products are labeled post PCR by using biotin-ddUTP/dUTP and terminal deoxynucleodityl transferase. A high-resolution confocal scanner is used to detect hybridization (68). The number of SNPs interrogated can be similar to Affymetrix GeneChip theoretically, but due to the PCR-based target preparation approach, Perlegen has screened up to 220 k SNPs (45). Pass rates were in the region of 90%.

Invader

Invader is the fourth technology used for WGAs (78) it is provided by Third Wave Technologies, and has been reviewed previously (48). In brief, the original method uses two oligonucleotides that can anneal to single strands of gDNA. One is the Invader oligonucleotide, which is complimentary to the sequence 3′ of the target SNP and ends with a nonmatching base overlapping the base at the SNP. The partner oligonucleotide is allele specific and extends toward the sequence 5′ of the SNP and contains additional nonoverlapping nucleotides at the 5′ prime. Annealing of the two oligonucleotides to the target DNA site leads to the formation of a three-dimensional structure that is recognized by the thermostable flap endonuclease (FEN), which cleaves the allele-specific oligonucleotide at the position of the SNP if the base at this position is complementary to the base of the gDNA. The allele-specific oligonucleotide contains a fluorescent label at its 5′ end that is thus released and able to generate fluorescence detected using a plate reader. The assay has been improved in the form of the serial invasive signal amplification reaction, which eliminates the need for target amplification by PCR and labeled allele-specific probes by utilizing a FRET cassette oligonucleotide (31). Invader combined with multiplexed PCR and a 384 well microfluidic card system to perform the reactions (InPlex) allowed this process to perform 100–200,000 assays (66).

CUSTOM ASSAY TECHNOLOGIES

Custom-designed SNP assays are utilized for replication and validation studies. These assays have to fulfill at least most of the following criteria: Allow rapid analysis of very high numbers of samples, have a high design to assay conversion rate for custom SNPs, and be robust and very cost-effective. Diverse and sophisticated molecular biological approaches have been incorporated into these technologies to fulfill these requirements.

Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF mass spectrometry): a tool for determining molecule mass, suitable for either protein or DNA and RNA analysis. Typically, single-stranded nucleic acid molecules at a range of 3–29 base pairs are resolved

Molecular Inversion Probes (MIP), (Affymetrix)

This method is based on padlock probes (33). These are 110–120-bp long probes, consisting of a middle part with sequences for PCR amplification and a 20-bp probe-specific tag, while the flanks contain locus- and SNP-specific sequences. The probes anneal immediately 5′ and 3′ of the SNP in question and become circular as a result of a gap fill step, using each of the four possible nucleotides in four separate reactions, followed by a ligation reaction. Probes that are not circularized are digested away by a combination of Exonuclease I and III, while the fully circularized probes are linearized and released from genomic DNA using Uracil-N-glycosilase. This means that in each of the four individual reactions, only probes containing a nucleotide complementary to one of the alleles in genomic DNA are left. The released probes are amplified by PCR, captured using a tag array, and labeled by hybridization with FAM-labeled Taqman probes complementary to sequences used for the PCR amplification step. Assays that utilize four colors, require just one tag array, and have increasing cost efficiency. Between 3,000–20,000 SNP assays can be performed in parallel. The four-color assay conversion rates are reported to be >80%, with pass rates of >98% and accuracy >99% (34).

The hardware required for MIP assays is similar to that used for Affymetrix Gene chips, apart from the requirement for a four-color scanner. The genotypes can be called automatically by a bespoke software that uses an expectation maximization (EM) algorithm (64).

iPlex Assays on the MassARRAY Platform (Sequenom)

These assays are based on matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry. iPlex is a single well reaction, in 384 well format, consisting of an initial multiplexed PCR step followed by a single-base primer extension reaction interrogating the SNP. Desalting occurs by the addition of anion exchange resin, and finally the reactions are arrayed onto chips containing 384 matrix spots. The chips are inserted into a MALDI-TOF instrument for analysis. Genotype calls are produced by determining the mass of the primer extension products which are designed to differ significantly and occupy distinct positions in the spectrum (39, 88). The iPlex assays can be multiplexed up to 40 plex and are very cost-effective at 12–40 multiplexing level. Key contributors to achieving consistently high levels of custom genotyping assays were the development of improved algorithms for PCR assay design combined with the single-base extension using acyclic bases and a proprietary polymerase. Pass rates of 95% or higher are possible, while accuracy is at >99%. Assay conversion rate is at a useful >80%. The MassARRAY system consists of a spotting instrument and a MALDI-TOF mass spectrometer, while typically one or more liquid handling robots and PCR blocks are required for high throughput (depending on MALDI-TOF instrument this can be 10 × 384 well reactions analyzed overnight). The main advantages of the system is that the assay requires unmodified oligonucleotide primers, which are easy to obtain at a low cost. Assay design software is provided in-house, allowing for high degrees of flexibility and customization; predesigned and validated assays are available at **http://www.realsnp.com/default.asp**. The system is also useful for a number of genomic applications such as methylation analysis or resequencing (74).

The Centaurus Assay (Nanogen)

The Centaurus genotyping method is based on the recently introduced endonuclease IV post-PCR genotyping system (47). First a standard PCR reaction is used to produce an amplicon containing the SNP of interest. An aliquot of the PCR reaction is used for the endonuclease reaction, which is an isothermal reaction at 50°–60°C that ideally takes place in a real-time PCR system. The SNP assay depends on the

combination of two allele-specific probes displaying a fluorescent dye at the 3′ end, a quencher at the 5′ end, and a short enhancer oligonucleotide (8–15 residues in length). The enhancer oligonucleotide hybridizes to the amplicon at 1-bp distance to the SNP site and 3′ relative to the allele-specific probes. The allele-specific probes distinguish between the two alleles at their 3′ end base, which carries the dye through a phosphodiester bond, mimicking an abasic site. This arrangement of the two probes on the target sequence is recognized by the endonuclease IV enzyme, which cleaves the dye only on the fully complementary probe, thus producing a fluorescent signal. The method has the advantage of using a short probe combination resulting in high-specificity and design flexibility.

SNPlex (Applied Biosystems)

The SNPlex assay is based on an oligonucleotide ligation/PCR assay (OLA/PCR) using allele-specific ZipCode™ probes and adaptors followed by hybridization of fluorescently labeled ZipChute™ probes, complementary to the ZipCode™ sequences. Up to 48 assays can be multiplexed. Detection of the genotypes is achieved by eluting the hybridized ZipChute™ probes and separating them by capillary electrophoresis (17, 93). Call rates are in the region of 97.5% and the accuracy above 99%, whereas the custom assay conversion rate is estimated to be lower than TaqMan (17). SNPlex is a multistep (eight steps) procedure designed for laboratories that use full automation, ideally in the form of two liquid handling robots and a plate wash station. One or more ABI 3730 capillary electrophoresis instruments provide high-sample throughput.

Goldengate and Infinium Assays (Illumina)

Infinium assays are also available as custom-designed products, called iSelect. iSelect assays are available for 6000–60,800 custom SNPs in a multisample (13) format. A minimum of 1152 samples is expected for a custom project. For a smaller number of SNPs, the Goldengate assay covers a spectrum of 96–1536 SNP assays, combining oligonucleotide ligation (OLA) and allele-specific extension reactions. The assay is a homogeneous reaction including biotinylation of the DNA and immobilization on avidin-coated particles, followed by annealing of three primers per SNP: one locus-specific and two allele-specific primers. This is followed by an extension reaction at the allele-specific oligonucleotide toward the locus-specific primer situated a few bases farther from the 3′ of it. A ligation reaction ligates the successfully extended allele-specific product to the locus-specific probe, a reaction that gives very high specificity to the assay. The probes have sequences allowing a PCR step with generic primers as well as sequences complementary to tags present in beads (67). The successfully extended and ligated products are amplified by PCR with fluorescently labeled primers. The PCR products are then denatured and hybridized to an array of beads (Sentrix Array) carrying sequences complementary to the locus-specific tags.

The system consists of a high-resolution scanner and computer workstation to detect fluorescent beads and decode the information used to generate the genotype calls. A fully automated Beadlab system is also available, albeit at a high cost. The assay conversion rate is in the region of 80%, while pass rates and accuracy are very high at >99% (84).

The TaqMan Assay and the OpenArray System (Applied Biosystems)

The Taqman assay is based on a 5′ nuclease assay (56), takes place in a single tube/well, and normally requires a real-time PCR machine such as the 7900 HT, by Applied Biosystems, to detect fluorescence. The method has two advantages: It requires just one simple reaction to be set up, while an ever-increasing number of prevalidated assays (currently at an impressive

Oligonucleotide ligation followed by PCR assay (OLA/PCR): ligation is a highly specific reaction, whereby only two DNA molecules perfectly annealed to the same template and with adjacent 5′ and 3′ ends are covalently linked. A PCR reaction is used to amplify the entire ligated DNA molecule

Single base extension assay (SBE): used to extend an oligonucleotide primer at the 3′ end by one base pair complementary to a template to which the primer is annealed. Very high specificity is achieved using thermostable enzymes

4.5 million for human and mouse) is available off the shelf by ABI. A web interface enables users to perform their own SNP assay designs and order.

An exciting development of this technology is the newly introduced OpenArray system, which uses plates that can accommodate 3072 individual reactions at the volume of 33 nl each in microscopic holes coated with a hydrophilic material. The plate is coated with a hydrophobic material and has the dimensions of a conventional microarray/microscope slide and contains 48 subarrays of 64 holes each. The system consists of a reagent dispenser that fills the holes with Taqman assay reagents and prepares the plates, PCR machines suitable for taking conventional microarray slides, and a scanner. Between 64 and 256 individual custom assays can be run per plate, at a rate of 48–12 samples, while 32 and 16 assays for up to 144 samples are in development. A major advantage of the system is that it can be completed within 8 h and is based on the reliable Taqman assays with a projected overall call rate of 99%. Up to 98,304 genotypes can be produced per day by a single system.

SNPstream (Beckman Coulter)

SNPstream combines a single base extension assay (SBE) and tag array technology. The first step is a multiplexed PCR reaction (up to 12 SNP-specific PCR assays) followed by a clean-up step and a primer extension reaction using tagged primers and labeled ddNTPs as terminators (Biodipy-Fluorescein and TAMRA). The products of the primer extension reaction are captured on a tag array, which is then scanned to detect the hybridized extension primers and produce calls (6). The multiplexed reactions are organized into six extension types, because only two types of fluorescently labeled terminators can be used in each multiplex (i.e., G/A, T/C, C/A,T/A, T/G, and C/G). Therefore, higher efficiencies are expected when a high number of SNP assays is required and allowed to be pooled in compatible groups. The assay conversion rate is claimed to be as high at 87.8%, pass rates at 98%, and accuracy at >99%. The assay requires, apart from thermal cyclers, a robotic plate handling system and a CCD scanner and has been used in clinical settings (18).

Other Technologies

Other methods are also available for SNP genotyping and have been well reviewed (for example, see 9, 48, 85). The most commonly applied methods are APEX (70), Dynamic Allele Specific Hybridization (DASH) (70), Molecular Beacons (60, 90), Primer Extension followed by MALDI-TOF (alternative to Sequenom's assays) (80), Pyrosequencing (24), and KASPar, a method based on competitive allele-specific PCR using FRET quencher cassette oligos, used by K-Biosciences (**http://www.kbioscience.co.uk/index.html**).

Methods for the Experimental Identification of Haplotypes

Haplotypes are combinations of alleles of genetic markers on a single chromosome. Haplotypes are important to understand functional effects of SNPs in *cis* and meiotic recombination events. Originally, it was possible to determine haplotypes over large genomic segments by genotyping either somatic cell hybrids (22) or sperm (38, 40, 99, 102). Experimental methods applicable to any genotype of DNA had been developed involving allele-specific PCR (61), MALDI-TOF genotyping on DNA dilutions (21), atomic force microscopy (101), and polony (clusters of molecules formed by PCR on a solid phase) PCR (62) but these had a limited range, usually less than 100 kb. Recently, the polony PCR approach has been modified and applied on arrays of stretched chromosome molecules, yielding impressive results over megabasepair-long regions and between chromosome arms (104). This method of haplotyping is more accurate than statistically inferred haplotypes. The method was also applied to detect chromosomal inversion events; another source of genomic variation (95).

Another intriguing approach to genotype haplotypes directly involves the capture by streptavidine magnetic beads of haploid chromosomal segments using biotinylated allele-specific probe hybridization (15). Further development of these methods and their application at high throughput will greatly support and validate the existing statistical approaches in the HapMap project as well as whole-genome sequencing. However, scaling up the polony haplotyping or the capturing method at a whole-genome level remains a challenge.

WHICH PLATFORM IS THE MOST POPULAR IN GWAs?

In evaluating 209 genome-wide association studies published until November 2008, one finds 103 studies using Affymetrix GeneChips, 83 using Infinium assays, 13 studies where data from both platforms have been generated either by direct runs (9) or by imputation (4), 8 studies by Perlegen, and two where Invader assays were used. More studies were based on Affymetrix GeneChips because the platform was already widely distributed in microarray facilities, and by the end of 2005, when a large number of studies such as the first Wellcome Trust Case Control were initiated, Affymetrix was already ahead in introducing chips with the highest marker content (500 k). Affymetrix has maintained the marker density lead, but now Illumina has become equally popular.

A number of studies have examined the relative advantages of the main genotyping platforms focusing mainly on marker content and global coverage of common variation at a genome-wide level (4, 69). Coverage is expressed as the fraction of common SNPs tagged by the SNPs on the chip based on HapMap data. More detailed studies have examined the variation in local coverage (53), or cost efficiency (52) also combined with imputation (3). In these studies the Illumina chips performed well and seem to be advantageous, at least when Caucasian populations are investigated. In particular, the HumanHap300 Beadchip is the most cost-effective option for Caucasian populations, as the genome coverage at $r^2 \geq 0.8$, achieved when imputed SNPs are included, reaches 81% compared to 73% for the Affymetrix SNP array 5.0 and 87% for the HumanHap550 (3).

Nevertheless, both platforms have had successful results, particularly when the studies were comparable in all other respects. For example, in the case of type II diabetes, two studies were performed using Affymetrix chips (81, 103) and two using Illumina chips (79, 86) and both generated highly comparable data.

Finally, other factors will influence the choice of SNP type, depending on the requirement of the study and whether rare cSNPs or genic SNPs are important and the ability to cover CNV, segmental duplication, and other areas of the unSNPable genome.

One factor to consider is the number of samples that can be processed simultaneously. Until very recently the very high density arrays, i.e., with 500,000 assays or more, were available in the form of one or even two chips per sample. Illumina has recently introduced multisample chips that allow the parallel processing of two samples with 1 million assays and four samples with 370,000, 610,000, and 660,000 assays. This development was complemented by the introduction of a faster scanning device and has significantly increased the throughput per system. In parallel, Affymetrix has been introducing high-throughput sample preparation formats (96 sample format) and chip handling automation for hybridization and scanning. As a result, the speed by which studies can be completed has now increased dramatically.

Examples of choices from the previously mentioned custom-designed assays in recently published GWA studies are iPlex (65, 75, 82, 96), SNPlex (7, 37), Centaurus (35, 44, 79), Taqman (32), and KASPar (19). These methods have also been applied to genotype a number of SNPs as sample identifiers prior to applying high-throughput approaches as described in Reference (91). Another interesting application of large-scale custom SNP assays has been presented by Keating et al. (41), whereby the iSelect Infinium assays have been used to

Table 1 Main characteristics of the most popular genotype assays applied in GWAs reviewed in terms of main molecular biology technique, relative assay, and sample throughput capability and detection technology basis

	TaqMan openarray	SNPlex	iPlex	Goldengate	Genechip	Infinium II
Assay type	5′exonuclease/PCR	OLA/PCR	Primer extension	Primer extension/ligation	Hybridization	Hybridization/Primer extension and ligation (I) or SBE (II)
Technology basis	TaqMan probes	Capillary electrophoresis	MALDI-TOF Mass spectrometry	Bead array	Oligonucleotide array	Bead array
Throughput/ person[a]	384–1536 samples/day	1536 samples/ 3 days	3840 samples/ 2.5 days	172 samples/3days	96 samples/ 5 days[with automation]	32–128 samples/ 5days[1–12 sample chips]
Multiplexing	64–256	24–48plex	12–40plex	384-1536	10,000–1.8 million	6000–1.2 million
Application	Medium custom SNP density Medium–large sample size	Medium custom SNP density Large sample size	Medium custom SNP density Large sample size	High custom or off the shelf SNP density Medium–large sample size	WGA studies; off the shelf SNP assays/ linkage analysis Small–large sample size	WGA studies/very high density custom SNP studies Small–large sample size

[a]The assay type and technology basis are included. A single person is able to operate all systems, and the expected sample throughput per person is indicated. Note that the figures are estimates and relate to basic versions of the equipment system and a normal up to 8-h work day, while throughputs will vary with degrees of automation and auxiliary equipment expansions. The complexity reduction step involves either locus-specific PCR reactions or genome fractionation by restriction enzyme digestion, whereas other technologies involve only signal amplification. (Modified from Reference 71.)

compile 50,000 assays to achieve high-density coverage of loci with high priority for cardiovascular disease. Such a chip is expected to complement commercial genotyping chips in GWA studies. For a summary of the features of currently popular methods in GWAs see **Table 1**.

CNV DETECTION

Recent research has shown that CNVs are likely to have functional consequences (5, 25, 89) and should be considered in GWAs (23). Traditionally, CNVs are detected using array CGH approaches (26), but since SNPs are used for GWAs, it makes sense to attempt to use SNP genotyping platforms for combined SNP and CNV analysis and thus to capture the bulk of human genomic variation. This prompted the development of algorithms that utilize SNP array data and aim to detect differences at every locus in copy number against the "expected" diploid reference genome. The measurements are based on the intensity of the allele-specific probes and the intensity ratios of the two alleles. Since both platforms (Affymetrix and Illumina) employ high SNP assay densities with a minimum average of one assay every 10 kb, if there is a change in copy number over a certain genomic segment all probe sets within it would be expected to behave similarly. For example, in case of a simple deletion, loss of heterozygosity and levels of intensity at the remaining allele-specific probe would be expected to be similar to the status of the heterozygote rather than of the homozygote. The goal is to identify such probe behavior out of the total noise generated by all SNP assays. Hidden Markov models (HMMs) are particularly suitable for this type of analysis by aiming to identify expected states (corresponding to homozygote or heterozygote, deletion, duplication, etc.) at defined genomic segments. HMMs were utilized for both the Illumina and Affymetrix platforms in the form of QuantiSNP (13), HMMseg (14, 16) and PennCNV (97, 98) for the Illumina Beadchips and PLASQ (49, 50), CNIT (54) and Birdsuite (46, 59) for Affymetrix GeneChips derived data. Other methods such as Trityper (27)

can be used to detect small deletions or null alleles at the single SNP assay level.

Systematic comparisons and validation of CNV data using a combination of different techniques are under way (43), including 1 Mi Beadchips (14) and Genechip 6.0 data (59) and show that mid-size deletions are detected by SNP platforms but duplications less well so. In addition, the largest part of human genomic variation appears to be caused by a finite number of common copy number polymorphisms (CNPs). This is encouraging for many reasons: as an ever-increasing number of individual human genomes are being sequenced, as in the 1000 genomes project (One thousand genomes project, **http://www.1000genomes.org/page.php**), all CNPs will be detected. As a result, existing platforms can be enriched with probes targeting these regions. Since the algorithms work, it is now possible to fill the gaps in the genome that have been left uncovered by probes, particularly to target at high-density, segmental duplications and also increase further the density within genic regions to increase the probability of detecting CNVs.

FUTURE DEVELOPMENTS

Two trends can be seen in WGA technologies: a move to higher numbers of markers (Affymetrix) and multisample arrays (Illumina). Increased numbers of assays will be required for CNV detection that may thus approach the sensitivity of tiled oligo arrays (discussed above), but this works against being able to accommodate higher numbers of samples on the current Affymetrix or Illumina chip formats.

In order to increase the speed and lower the cost associated with performing WGA studies, new formats will be required that enable the processing of 96 or even 384 samples in parallel. Affymetrix is already introducing a new 96 plate format system (GeneChip HT PM Array Plate) containing miniaturized arrays of chips, currently suitable for gene expression studies. These arrays can be used in an automated microarray processing system introduced by Affymetrix, the GeneTitan. It is conceivable that such a system could be adopted for array-based genotyping assays and will lead to increased throughput and decreased costs by reducing hands-on and overall processing time.

For assays applied in replication and validation studies, further cost reductions and sample throughput increases will be needed. These can be achieved by reducing the reaction volumes and introducing higher-density plate, chip, or array formats. The successful move to using microfluidic devices for performing nanoliter-scale PCR reactions by ABI and Third Wave Technologies can be followed by the application of even more sophisticated microfluidic devices allowing multiplexed PCR products to be generated. For example, such devices are produced by RainDance Technologies (Lexington, MA, USA) and allow up to 1000s of individual PCR reactions to take place in parallel from a single sample. Primer extension reactions should also be possible with this method. A variety of methods such as MALDI-TOF can subsequently be applied to analyze the products.

Ultimately, dramatic cost savings will be achieved by applying radically different technologies, similar to the breakthroughs that have been achieved in next-generation sequencing. One of the technologies that stands out is nanopore technology that has the potential to form the basis of "third generation" sequencing instruments (8). The technology could be used for low-cost, targeted SNP genotyping, as it may be possible to use optical detection of fluorescent probes annealed to target sequences at nanopores (10) or to perform "minisequencing" of amplified DNA fragments similar to the sequencing approaches reviewed in Reference (8).

CONCLUSIONS

The current high-throughput genotyping platforms, whether GeneChip or Beadchip based, are able to perform robustly enough for laboratories worldwide to successfully complete GWA studies. The Affymetrix system currently offers chips at considerably lower

Copy number polymorphisms (CNPs): CNVs with a frequency higher than 1% in the population

cost per marker than Illumina (for example, the Affymetrix 6.0 GeneChip contains 1.8 Mi assays and costs half as much as Illumina's 1.2 million assay chip). The one most appropriate to use will depend on considerations such as whether a particular platform is already installed, the population to be studied, and the previous experience of the investigators. The more SNP assays are included in the future resulting in very dense coverage, the more likely that imputation will blur the differences between the two platforms. What may become the distinguishing feature is the additional content aimed at enabling CNP and CNV detection.

One additional issue is the ability of these systems to allow the user to design and incorporate assays for individual SNPs of choice. These SNPs can be population-specific or disease-specific, as exemplified in Reference (41). In general, a huge gap in cost (in the region of 100- to 1000-fold) per genotype still exists between the high-throughput platforms and the custom low-throughput platforms and needs to be reduced. Over the short term, the gap could narrow through the application of microfluidic devices that allow reactions to take place at a nanoliter scale, similar to the OpenArray system. Additional overall cost savings could be achieved through the availability of validated assays for the largest proportion of the ~10 million SNPs in the human genome. This will reduce waste and unnecessary duplication of assay validation efforts.

Will the high-throughput genotyping platforms be obsolete in the near future and replaced by low-cost DNA sequencing? As we move toward ever more cost-effective sequencing technology and the $1000 genome (10), an increasing number of studies will likely employ initially region-specific sequencing to detect rare variants. Once the $1000 genome becomes reality, whole-genome sequencing may be expected to replace high-throughput genotyping. However, once causative variants have been detected, genotyping platforms will be increasingly used in the clinic (72).

SUMMARY POINTS

1. Current genotyping technology offers a variety of possibilities for genetic research. It is possible to assay a sample with a wide spectrum of multiplexed assays, from tens to millions of markers at the same time.
2. The current high-throughput genotyping platforms, whether based on the Affymetrix GeneChip or Illumina's Beadchip, are able to deliver genome-wide coverage and a robust performance that allows laboratories worldwide to successfully complete GWA studies.
3. Although it is possible to detect CNVs with the current SNP arrays, it is important to increase the marker density further to cover CNVs, CNPs, and other structural variations.
4. The low-assay throughput technologies are robust and deliver high-sample throughput and play a key role in replication studies.
5. Methods for long-range haplotyping based on polony PCR and direct haplotype capture have been developed, and these can be used to identify recombination events and structural variation.
6. Further technological advances are required and are possible in order to achieve lower costs and even higher throughput. Developments in microfluidics and nanotechnologies should facilitate these advances in the near future.

FUTURE ISSUES

1. Sample throughput in ultrahigh multiplex assays must be increased;
2. Marker content in CNP, CNV, and unSNPable segments must be increased further;
3. Costs of low- to mid-throughput assays must be compressed toward 1 pence or cent/genotype;
4. The number of validated, off-the-shelf assays must be increased;
5. Platforms must utilize technological advances in high-throughput sequencing.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I would like to thank Ghazala Mirza for providing the material for **Figures 1** and **2**, Silvia Paracchini for helpful comments on the manuscript and Helen Butler for editing.

LITERATURE CITED

1. **2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–78**
2. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. 2005. A haplotype map of the human genome. *Nature* 437:1299–320
3. Anderson CA, Pettersson FH, Barrett JC, Zhuang JJ, Ragoussis J, et al. 2008. Evaluating the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms. *Am. J. Hum. Genet.* 83:112–19
4. Barrett JC, Cardon LR. 2006. Evaluating coverage of genome-wide association studies. *Nat. Genet.* 38:659–62
5. Beckmann JS, Estivill X, Antonarakis SE. 2007. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat. Rev. Genet.* 8:639–46
6. Bell PA, Chaturvedi S, Gelfand CA, Huang CY, Kochersperger M, et al. 2002. SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *Biotechniques* (Suppl.):70–77
7. Bouatia-Naji N, Rocheleau G, Van Lommel L, Lemaire K, Schuit F, et al. 2008. A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels. *Science* 320:1085–88
8. **Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, et al. 2008. The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* 26:1146–53**
9. Breen G. 2002. Novel and alternate SNP and genetic technologies. *Psychiatr. Genet.* 12:83–88
10. Cantor CR. 2008. Quantitative, sensitive analysis of DNA and RNA. *Nucleic Acids Symp. Ser.* 52:15
11. Chan EY. 2005. Advances in sequencing technology. *Mutat. Res.* 573:13–40
12. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, et al. 2007. Replicating genotype-phenotype associations. *Nature* 447:655–60
13. Colella S, Yau C, Taylor JM, Mirza G, Butler H, et al. 2007. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.* 35:2013–25
14. **Cooper GM, Zerr T, Kidd JM, Eichler EE, Nickerson DA. 2008. Systematic assessment of copy number variant detection via genome-wide SNP genotyping. *Nat. Genet.* 40:1199–203**

1. This study by the Wellcome Trust Case Control Consortium was the largest genome-wide association of its time using common controls for seven diseases.

8. This review provides a comprehensive description of current approaches in nanopore based DNA sequencing.

14. A systematic comparison of CNV detection across SNP genotyping platforms.

15. Dapprich J, Ferriola D, Magira EE, Kunkel M, Monos D. 2008. SNP-specific extraction of haplotype-resolved targeted genomic regions. *Nucleic Acids Res.* 36:e94
16. Day N, Hemmaplardh A, Thurman RE, Stamatoyannopoulos JA, Noble WS. 2007. Unsupervised segmentation of continuous genomic data. *Bioinformatics* 23:1424–26
17. De la Vega FM, Lazaruk KD, Rhodes MD, Wenz MH. 2005. Assessment of two flexible and compatible SNP genotyping platforms: TaqMan SNP genotyping assays and the SNPlex genotyping system. *Mutat. Res.* 573:111–35
18. Denomme GA, Van Oene M. 2005. High-throughput multiplex single-nucleotide polymorphism analysis for red cell and platelet antigen genotypes. *Transfusion* 45:660–66
19. Di Bernardo MC, Crowther-Swanepoel D, Broderick P, Webb E, Sellick G, et al. 2008. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nat. Genet.* 40:1204–10
20. Di X, Matsuzaki H, Webster TA, Hubbell E, Liu G, et al. 2005. Dynamic model based algorithms for screening and genotyping over 100 K SNPs on oligonucleotide microarrays. *Bioinformatics* 21:1958–63
21. Ding C, Cantor CR. 2003. Direct molecular haplotyping of long-range genomic DNA with M1-PCR. *Proc. Natl. Acad. Sci. USA* 100:7449–53
22. Douglas JA, Boehnke M, Gillanders E, Trent JM, Gruber SB. 2001. Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. *Nat. Genet.* 28:361–64
23. Estivill X, Armengol L. 2007. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet.* 3:1787–99
24. Fakhrai-Rad H, Pourmand N, Ronaghi M. 2002. Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. *Hum. Mutat.* 19:479–85
25. Feuk L, Carson AR, Scherer SW. 2006. Structural variation in the human genome. *Nat. Rev. Genet.* 7:85–97
26. Fiegler H, Redon R, Andrews D, Scott C, Andrews R, et al. 2006. Accurate and reliable high-throughput detection of copy number variation in the human genome. *Genome Res.* 16:1566–74
27. Franke L, de Kovel CG, Aulchenko YS, Trynka G, Zhernakova A, et al. 2008. Detection, imputation, and association analysis of small deletions and null alleles on oligonucleotide arrays. *Am. J. Hum. Genet.* 82:1316–33
28. **Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–61**
29. Giannoulatou E, Yau C, Colella S, Ragoussis J, Holmes CC. 2008. GenoSNP: a variational Bayes within-sample SNP genotyping algorithm that does not require a reference population. *Bioinformatics* 24:2209–14
30. Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS. 2005. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat. Genet.* 37:549–54
31. Hall JG, Eis PS, Law SM, Reynaldo LP, Prudent JR, et al. 2000. Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. *Proc. Natl. Acad. Sci. USA* 97:8272–77
32. Han J, Kraft P, Nan H, Guo Q, Chen C, et al. 2008. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* 4:e1000074
33. Hardenbol P, Baner J, Jain M, Nilsson M, Namsaraev EA, et al. 2003. Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat. Biotechnol.* 21:673–78
34. Hardenbol P, Yu F, Belmont J, Mackenzie J, Bruckner C, et al. 2005. Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res.* 15:269–75
35. Helgadottir A, Thorleifsson G, Magnusson KP, Gretarsdottir S, Steinthorsdottir V, et al. 2008. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat. Genet.* 40:217–24
36. Hinds DA, Stokowski RP, Patil N, Konvicka K, Kershenobich D, et al. 2004. Matching strategies for genetic association studies in structured populations. *Am. J. Hum. Genet.* 74:317–25
37. Hofmann S, Franke A, Fischer A, Jacobs G, Nothnagel M, et al. 2008. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat. Genet.* 40:1103–6

28. This publication describes the most recent version of HapMap (HapMAp II). It serves as the basis for genotype imputation, SNP assay design, and SNP haplotype analysis.

38. Jeffreys AJ, Neumann R, Panayi M, Myers S, Donnelly P. 2005. Human recombination hot spots hidden in regions of strong marker association. *Nat. Genet.* 37:601–6
39. Jurinke C, Van Den Boom D, Cantor CR, Koster H. 2002. Automated genotyping using the DNA MassARRAY technology. *Methods Mol. Biol.* 187:179–92
40. Kauppi L, Jeffreys AJ, Keeney S. 2004. Where the crossovers are: recombination distributions in mammals. *Nat. Rev. Genet.* 5:413–24
41. Keating BJ, Tischfield S, Murray SS, Bhangale T, Price TS, et al. 2008. Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS ONE* 3:e3583
42. Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, et al. 2003. Large-scale genotyping of complex DNA. *Nat. Biotechnol.* 21:1233–37
43. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, et al. 2008. Mapping and sequencing of structural variation from eight human genomes. *Nature* 453:56–64
44. Kiemeney LA, Thorlacius S, Sulem P, Geller F, Aben KK, et al. 2008. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat. Genet.* 40:1307–12
45. Kooner JS, Chambers JC, Aguilar-Salinas CA, Hinds DA, Hyde CL, et al. 2008. Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat. Genet.* 40:149–51
46. Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, et al. 2008. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* 40:1253–60
47. Kutyavin IV, Milesi D, Belousov Y, Podyminogin M, Vorobiev A, et al. 2006. A novel endonuclease IV post-PCR genotyping system. *Nucleic Acids Res.* 34:e128
48. Kwok PY. 2001. Methods for genotyping single nucleotide polymorphisms. *Annu. Rev. Genomics Hum. Genet.* 2:235–58
49. Laframboise T, Harrington D, Weir BA. 2007. PLASQ: a generalized linear model-based procedure to determine allelic dosage in cancer cells from SNP array data. *Biostatistics* 8(2):323–36
50. Laframboise T, Weir BA, Zhao X, Beroukhim R, Li C, et al. 2005. Allele-specific amplification in cancer revealed by SNP array analysis. *PLoS Comput. Biol.* 1:e65
51. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
52. Li C, Li M, Long JR, Cai Q, Zheng W. 2008. Evaluating cost efficiency of SNP chips in genome-wide association studies. *Genet. Epidemiol.* 32:387–95
53. Li M, Li C, Guan W. 2008. Evaluation of coverage variation of SNP chips for genome-wide association studies. *Eur. J. Hum. Genet.* 16:635–43
54. Lin CH, Huang MC, Li LH, Wu JY, Chen YT, Fann CS. 2008. Genome-wide copy number analysis using copy number inferring tool (CNIT) and DNA pooling. *Hum. Mutat.* 29:1055–62
55. Lips EH, Dierssen JW, van Eijk R, Oosting J, Eilers PH, et al. 2005. Reliable high-throughput genotyping and loss-of-heterozygosity detection in formalin-fixed, paraffin-embedded tumors using single nucleotide polymorphism arrays. *Cancer Res.* 65:10188–91
56. Livak KJ. 2003. SNP genotyping by the 5′-nuclease reaction. *Methods Mol. Biol.* 212:129–47
57. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, et al. 2008. Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* 82:477–88
58. Matsuzaki H, Dong S, Loi H, Di X, Liu G, et al. 2004. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat. Methods* 1:109–11
59. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, et al. 2008. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* 40:1166–74
60. Mhlanga MM, Malmberg L. 2001. Using molecular beacons to detect single-nucleotide polymorphisms with real-time PCR. *Methods* 25:463–71
61. Michalatos-Beloin S, Tishkoff SA, Bentley KL, Kidd KK, Ruano G. 1996. Molecular haplotyping of genetic markers 10 kb apart by allele-specific long-range PCR. *Nucleic Acids Res.* 24:4841–43
62. Mitra RD, Butty VL, Shendure J, Williams BR, Housman DE, Church GM. 2003. Digital genotyping and haplotyping with polymerase colonies. *Proc. Natl. Acad. Sci. USA* 100:5926–31

46. Describes algorithms that integrate existing CNP information, CNV detection, SNP calling, and association analysis.

63. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, et al. 2007. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 448:470–73
64. Moorhead M, Hardenbol P, Siddiqui F, Falkowski M, Bruckner C, et al. 2006. Optimal genotype determination in highly multiplexed SNP data. *Eur. J. Hum. Genet.* 14:207–15
65. O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, et al. 2008. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat. Genet.* 40:1053–55
66. Ohnishi Y, Tanaka T, Ozaki K, Yamada R, Suzuki H, Nakamura Y. 2001. A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.* 46:471–77
67. Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. 2002. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* (Suppl.):56–61
68. Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, et al. 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 294:1719–23
69. Pe'er I, de Bakker PI, Maller J, Yelensky R, Altshuler D, Daly MJ. 2006. Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat. Genet.* 38:663–67
70. Podder M, Ruan J, Tripp BW, Chu ZE, Tebbutt SJ. 2008. Robust SNP genotyping by multiplex PCR and arrayed primer extension. *BMC Med. Genomics* 1:5
71. Prince JA, Feuk L, Howell WM, Jobs M, Emahazion T, et al. 2001. Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. *Genome Res.* 11:152–62
72. Ragoussis J. 2006. Genotyping technologies for all. *Drug Discov. Today Technol.* 3:115–22
73. Ragoussis J, Elvidge G. 2006. Affymetrix GeneChip system: moving from research to the clinic. *Expert. Rev. Mol. Diagn.* 6:145–52
74. Ragoussis J, Elvidge GP, Kaur K, Colella S. 2006. Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry in genomics research. *PLoS Genet.* 2:e100
75. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, et al. 2008. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nat. Genet.* 40:1216–23
76. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. 2006. Global variation in copy number in the human genome. *Nature* 444:444–54
77. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, et al. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928–33
78. Sakamoto H, Yoshimura K, Saeki N, Katai H, Shimoda T, et al. 2008. Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat. Genet.* 40:730–40
79. Salonen JT, Uimari P, Aalto JM, Pirskanen M, Kaikkonen J, et al. 2007. Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. *Am. J. Hum. Genet.* 81:338–45
80. Sauer S, Lechner D, Berlin K, Plancon C, Heuermann A, et al. 2000. Full flexibility genotyping of single nucleotide polymorphisms by the GOOD assay. *Nucleic Acids Res.* 28:E100
81. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, et al. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–36
82. Schormair B, Kemlink D, Roeske D, Eckstein G, Xiong L, et al. 2008. PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. *Nat. Genet.* 40:946–48
83. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, et al. 2007. Strong association of de novo copy number mutations with autism. *Science* 316:445–49
84. Shen R, Fan JB, Campbell D, Chang W, Chen J, et al. 2005. High-throughput SNP genotyping on universal bead arrays. *Mutat. Res.* 573:70–82
85. Shi MM. 2001. Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clin. Chem.* 47:164–72
86. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, et al. 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881–85
87. Steemers FJ, Chang W, Lee G, Barker DL, Shen R, Gunderson KL. 2006. Whole-genome genotyping with the single-base extension assay. *Nat. Methods* 3:31–33
88. Storm N, Darnhofer-Patel B, Van Den Boom D, Rodi CP. 2003. MALDI-TOF mass spectrometry-based SNP genotyping. *Methods Mol. Biol.* 212:241–62

89. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, et al. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315:848–53
90. Tapp I, Malmberg L, Rennel E, Wik M, Syvanen AC. 2000. Homogeneous scoring of single-nucleotide polymorphisms: comparison of the 5′-nuclease TaqMan assay and Molecular Beacon probes. *Biotechniques* 28:732–38
91. Teo YY, Inouye M, Small KS, Fry AE, Potter SC, et al. 2008. Whole genome-amplified DNA: insights and imputation. *Nat. Methods* 5:279–80
92. Teo YY, Inouye M, Small KS, Gwilliam R, Deloukas P, et al. 2007. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* 23:2741–46
93. Tobler AR, Short S, Andersen MR, Paner TM, Briggs JC, et al. 2005. The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. *J. Biomol. Tech.* 16:398–406
94. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, et al. 2007. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat. Genet.* 39:857–64
95. Turner DJ, Shendure J, Porreca G, Church G, Green P, et al. 2006. Assaying chromosomal inversions by single-molecule haplotyping. *Nat. Methods* 3:439–45
96. Van Den Oord EJ, Kuo PH, Hartmann AM, Webb BT, Moller HJ, et al. 2008. Genome-wide association analysis followed by a replication study implicates a novel candidate gene for neuroticism. *Arch. Gen. Psychiatry* 65:1062–71
97. Wang K, Chen Z, Tadesse MG, Glessner J, Grant SF, et al. 2008. Modeling genetic inheritance of copy number variations. *Nucleic Acids Res.* 36:e138
98. Wang K, Li M, Hadley D, Liu R, Glessner J, et al. 2007. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 17:1665–74
99. Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, et al. 2005. Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* 308:107–11
100. Wirtenberger M, Hemminki K, Burwinkel B. 2006. Identification of frequent chromosome copy-number polymorphisms by use of high-resolution single-nucleotide-polymorphism arrays. *Am. J. Hum. Genet.* 78:520–22
101. Woolley AT, Guillemette C, Li Cheung C, Housman DE, Lieber CM. 2000. Direct haplotyping of kilobase-size DNA using carbon nanotube probes. *Nat. Biotechnol.* 18:760–63
102. Yauk CL, Bois PR, Jeffreys AJ. 2003. High-resolution sperm typing of meiotic recombination in the mouse MHC Ebeta gene. *EMBO J.* 22:1389–97
103. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, et al. 2007. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316:1336–41
104. Zhang K, Zhu J, Shendure J, Porreca GJ, Aach JD, et al. 2006. Long-range polony haplotyping of individual human chromosome molecules. *Nat. Genet.* 38:382–87

89. This paper describes for the first time the association between SNPs, CNVs, and levels of gene expression, identifying *cis* and *trans* effects.

Annual Review of Genomics and Human Genetics

Volume 10, 2009

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Genomics and Human Genetics* articles may be found at http://genom.annualreviews.org/